A Spatio-Temporal Ageing Atlas of the Proximal Femur

Mohsen Farzi, *Member, IEEE*, Jose M. Pozo, Eugene McCloskey, Richard Eastell, Nicholas Harvey, J. Mark Wilkinson, and Alejandro F. Frangi, *Fellow, IEEE*

***Abstract*— Osteoporosis is an age-associated disease characterised by low bone mineral density (BMD) and micro-architectural deterioration leading to enhanced fracture risk. Conventional dual-energy X-ray absorptiometry (DXA) analysis has facilitated our understanding of BMD reduction in specific regions of interest (ROIs) within the femur, but cannot resolve spatial BMD patterns nor reflect age-related changes in bone microarchitecture due to its inherent averaging of pixel BMD values into large ROIs. To address these limitations and develop a comprehensive model of involutional bone loss, this paper presents a fully automatic pipeline to build a spatio-temporal atlas of ageing bone in the proximal femur. A new technique, termed DXA region free analysis (DXA RFA), is proposed to eliminate morphological variation between DXA scans by warping each image into a reference template. To construct the atlas, we use unprocessed DXA data from Caucasian women aged 20-97 years participating in three cohort studies in Western Europe (n >13,000). A novel calibration procedure, termed *quantile matching regression*, is proposed to integrate data from different DXA manufacturers. Pixel-wise BMD evolution with ageing was modelled using smooth quantile curves. This technique enables characterisation of spatially-complex BMD change patterns with ageing, visualised using heat-maps. Furthermore, quantile curves plotted at different pixel coordinates showed consistently different rates of bone loss at different regions within the femoral neck. Given the close relationship between spatio-temporal bone loss and osteoporotic fracture, improved understanding of the bone ageing process could lead to enhanced prognostic, preventive and therapeutic strategies for the disease.**

***Index Terms*— Spatio-temporal atlas, DXA, region free analysis, osteoporosis.**

Manuscript received August 12, 2019; revised September 11, 2019; accepted September 22, 2019. Date of publication October 23, 2019; date of current version April 30, 2020. The work of M. Farzi was supported by the Ph.D. Fellowship from the Medical Research Council/Arthritis Research U.K. Centre for Integrated Research into Musculoskeletal Ageing under Grant MR/P020941/1. This work was supported in part by the European Commission through BackUP Project under Grant H2020-SC1-PM-17-2017-777090 and in part by the Engineering and Physical Sciences Research Council through the MedIAN Network under Grant EP/N026993/1. *(J. Mark Wilkinson and Alejandro F. Frangi contributed equally to this work.) (Corresponding author: J. Mark Wilkinson.)*

M. Farzi is with the Department of Oncology and Metabolism, The University of Sheffield, Sheffield S10 2RX, U.K., and also with the Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), The University of Leeds, Leeds LS2 9JT, U.K. (e-mail: m.farzi@leeds.ac.uk).

J. M. Pozo and A. F. Frangi are with the Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), The University of Leeds, Leeds LS2 9JT, U.K. (e-mail: j.m.pozo@leeds.ac.uk; a.frangi@leeds.ac.uk).

E. McCloskey, R. Eastell, and J. M. Wilkinson are with the Department of Oncology and Metabolism, The University of Sheffield, Sheffield S10 2RX, U.K., and also with the MRC-Arthritis Research UK Centre for Integrated research into Musculoskeletal Ageing (CIMA), The University of Sheffield, Sheffield S10 2RX, U.K. (e-mail: e.v.mccloskey@sheffield.ac.uk; r.eastell@sheffield.ac.uk; j.m.wilkinson@sheffield.ac.uk).

N. Harvey is with the MRC Lifecourse Epidemiology Unit, The University of Southampton, Southampton SO16 6YD, U.K. (e-mail: nch@mrc.soton.ac.uk).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2019.2945219

I. Introduction

AGEING is associated with a gradual and progressive bone loss, which predisposes to osteoporosis. Osteoporosis is a bone disease characterised by low bone mass and micro-architectural deterioration, and improving the understanding of the bone ageing process interests the osteoporosis research community [1], [2]. To facilitate this understanding, we propose a method to develop a spatio-temporal atlas of ageing bone in the femur.

Spatio-temporal imaging atlases quantitatively describe detailed anatomical or functional phenotypes from large imaging studies. For example, brain atlases enable quantitation of disease progression in Alzheimer's Disease [3]. However, to the best of our knowledge, no spatio-temporal atlas of ageing bone has been developed in osteoporosis research. Developing a comprehensive model of involutional bone loss is a challenging task. First, this requires a robust and accurate quantification technique for bone mineral density (BMD) measurement and its spatial distribution. Dual-energy X-ray Absorptiometry (DXA) is the reference gold standard to measure BMD in clinical practice [4]. In conventional DXA analysis, BMD data is acquired at the individual pixel level but at the analysis stage these values are averaged in *a priori* specified regions of interest (ROIs) to compensate for shape variation between scans (Fig. 1). Data averaging reduces the ability of the technique to quantitate local variation in textural BMD patterns of clinical relevance to disease progression with ageing.

The second challenge is the ability to homogenise BMD measurements across different technologies, as systematic differences in instrument calibration exist between different proprietary DXA manufacturers [5]–[7]. Two broad cross-calibration procedures are commonly used. In one approach, each scanner is separately calibrated by fitting bone phantom measurements to its nominal density values. Pearson *et al.* [8] suggested an exponential curve and explored the technique using the European Spine Phantom (ESP) prototype. In the

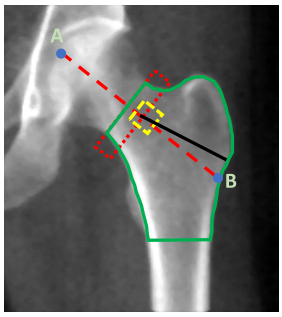


Fig. 1. Femoral Regions of Interest (ROIs). The green solid contour around the bone shows the total hip. The neck and the wards' triangle regions are identified with red and yellow rectangles, respectively. The A-B line shows the hip axis and the black line delineates between the trochanteric and the inter-trochanteric regions.

other approach, different scanners are calibrated simultaneously using density values measured on a common group of individuals [5]–[7].

Both DXA calibration procedures have several key limitations. Cross-calibration using phantom measurements is challenged by a study conducted under the auspices of the International DXA Standardisation Committee (IDSC) [5]. Genant *et al.* [5] showed a disagreement between regression curves fitted to the phantom measurements and those fitted to the human measurements. The second approach requires repeated measurements of each subject across all machines [5]–[7]. This can be a seriously limiting factor in large multi-centre studies, where the first approach may be preferred in practice [9].

This paper offers solutions to these challenges: To maintain fidelity to high-resolution pixel BMD values, we have previously proposed a *region free analysis* (RFA) approach and demonstrated its application to analysing periprosthetic BMD changes for patients who received a hip prosthesis [10]–[12]. DXA RFA aligns each individual scan to a reference template and so eliminates the morphological variation between scans. This deformable image alignment establishes a virtual correspondence between pixel coordinates enabling statistical inference at the pixel level. To control the correspondence between scans, the initial RFA technique used a set of anatomical landmark points selected semi-automatically around the prosthesis and the bone contour. Here, we extend the technique to the native femur and propose a fully automatic formulation applicable to large-scale datasets (section II-B).

To amalgamate data from different centres, we propose a novel calibration procedure termed quantile matching regression. The proposed technique uses BMD measurements collected from individuals but the requirement for scanning the same group of subjects on all the machines is moderated. In this method, different subject groups with similar geographic, ethnic and demographic characteristics scanned on each machine are assumed to be independent and identically distributed samples from the same population. The cross-calibration is achieved by matching the distribution of the BMD values over the matched sample groups across the centres.

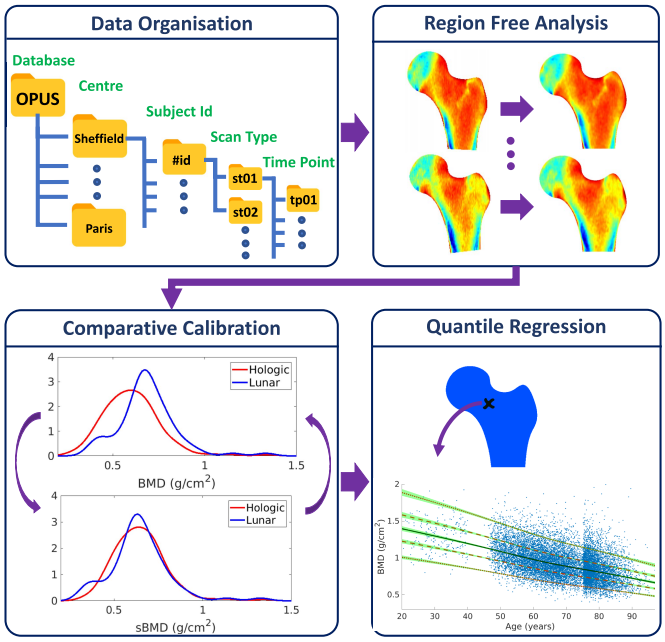


Fig. 2. Bone ageing analysis pipeline. Scans are automatically organised into sub-folders according to the study ID, geographic location, subject ID, anatomic site, and follow-up time points. Each scan is then warped into a reference domain to eliminate morphological variations. Pixel BMD values are calibrated across different centres so the probability density functions match one another for a subset of samples matched for gender, age, body mass index, ethnicity, scan side, and geographic location. Finally, a set of smooth quantile curves is fitted to the standardised pixel BMD values for each pixel coordinate.

This paper aims to develop the first spatio-temporal reference atlas of ageing bone in the femur using DXA data from over 13,000 subjects (Fig. 2). To model the temporal BMD evolution as a function of age, quantile curves were fitted using vector generalised additive models (VGAMs) per each pixel coordinate. Preliminary results using a subset of $n = 1,714$ subjects were presented in [13]. This paper expands our previous work: first, the mathematical details are provided here and the method is also extended to allow contribution of confounding variables such as body mass index (BMI) in the developed atlas. Second, the method is applied to a considerably larger dataset here. This increase in sample size reduces the uncertainty around each quantile ageing trajectory leading to a precise and accurate model. Third, extensive validation using experimental data was performed to evaluate each module in the proposed pipeline (Fig. 2). Fourth, given the atlas is developed on cross-sectional data, the ability of the atlas to predict longitudinal changes is validated using a subset of data ($n = 120$) with actual BMD measurements at baseline and six years later. The proposed atlas is the first comprehensive spatio-temporal model of ageing bone. The observed spatially-complex ageing patterns may be used to better map the development of osteoporosis with ageing and enhance the prediction of consequent fragility fractures.

II. Methods

Fig. 2 shows the conceptual outline of the proposed method. Below, different steps of the proposed framework are explained

in detail: pre-processing and data organisation, region free analysis, comparative calibration, and quantile regression.

A. Pre-Processing and Data Organisation

The raw data from the densitometer is not immediately usable for analysing BMD maps. We used Hologic Apex v3.2 (Hologic Inc, Waltham, MA) and Lunar enCORE v16 (GE Healthcare, Madison, WI) proprietary software to extract pixel BMD information for scans collected on a Hologic QDR 4500A or a Lunar iDXA densitometer, respectively.

Spatial resolution and signal-to-noise ratio (SNR) vary between the 2 densitometer manufacturers. For example, the spatial resolution, expressed as height×width, is 0.50×0.90 mm^2 for a Hologic QDR 4500A scanner and 0.25×0.30 mm^2 for a Lunar iDXA scanner. Based on our measurements, while the Lunar system provides a better resolution by a factor of two in height and three in width, pixel-wise SNR is ca. 10 dB higher in the Hologic system, as estimated using a subset of repeated DXA measurements. To enable data integration, an appropriate analysis scale should be selected so both the spatial resolution and the pixel-wise SNR are consistent across the two systems. For selection of an appropriate scale, all scans were resampled at an isotropic spatial resolution of 0.5×0.5 mm^2. Following resampling, each image was smoothed with a Gaussian kernel to enhance the SNR. Given the higher SNR for the Hologic system, $\sigma = 0$ and $\sigma = 4.5$ were selected for the Hologic and the Lunar systems so both systems have the same SNR of 22.4 dB.

B. Region Free Analysis

RFA aims to find a set of coordinate transformations such that the warped scans are aligned with each other in the template domain. Therefore, each pixel coordinate in the template domain corresponds to the same anatomical location in the image domain. This correspondence allows pixel-level inference of the BMD values. The RFA technique has three steps (Fig. 3): automatic landmark extraction, template derivation, and pairwise registration between the reference template and each scan.

1) Automatic Landmark Extraction: This section addresses the problem of automatically locating prominent feature points in the femur. These feature points are used to compute the geometrical warp between the image domain and the template (see section II-B.3). A standard approach to this problem is to first build a model of shape and texture variation from a manually labelled training set, and then fit the model to an unseen image [14]. For improved localisation accuracy, here we used constrained local model (CLM) algorithm, which combines the flexibility of appearance models with global shape constraints [15].

In CLMs, a joint shape and texture model is learned in a similar manner to Statistical Appearance Models (SAMs) [16]; however, the texture sampling method is in the form of rectangle patches around landmark points. In the CLM framework, a *response image* is generated per each landmark point independently. To generate a response image for the m^{th} landmark point, random patches at its local neighbourhood are selected

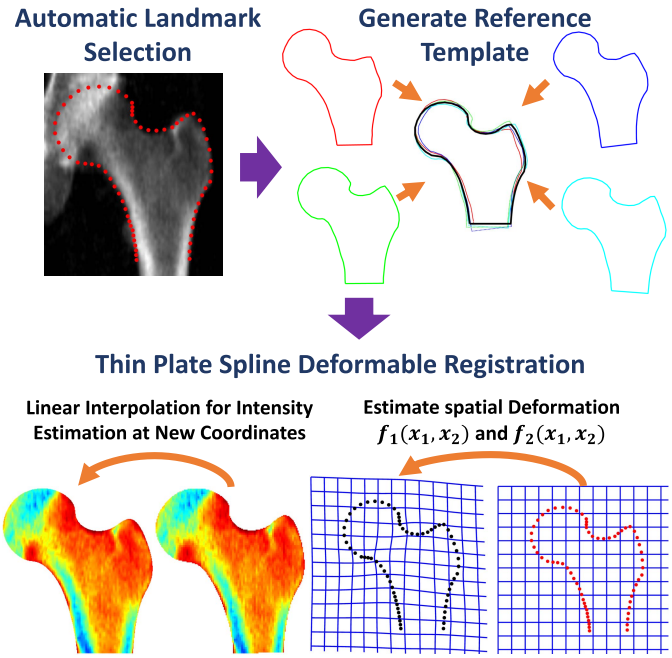


Fig. 3. Conceptual illustration of region free analysis. Sixty-five landmark points are automatically selected around the bone contour. A reference shape is learned by averaging over all the scans after being aligned to a common position, scale, and orientation. A thin plate spline (TPS) deformation function is fitted for each individual scan such that the controlling landmark points are mapped to the corresponding reference landmark points in the template. Given the warp in the space, pixel intensities are estimated using a linear interpolation technique.

and the correlation of each patch with *a priori* trained template is computed. Then, the objective function $\mathcal{J}(\boldsymbol{b}_s)$ is maximised to find the optimal shape parameters [15].

$$\mathcal{J}(\boldsymbol{b}_s) = \alpha \sum_{m=1}^{M} R_m(x'_{1,m}, x'_{2,m}) - \sum_{j=1}^{J} \frac{b_j^2}{\lambda_j}, \tag{1}$$

where $[x'_{1,m}, x'_{2,m}]^T$ is the current estimation of landmark point m, R_m is the response image for point m, J is the total number of shape parameters, and λ_j are the corresponding eigenvalues of the shape model. The algorithm iterates until convergence happens.

Lindner *et al.* [17] applied CLMs in the setting of femur segmentation. However, instead of computing the correlation with a template, random forest voting was deployed to generate the response images where the decision trees voted for the required displacements. To initialise the landmark points, a Hough-like approach was utilised to automatically detect the femur in the scan [18]. Here, we deployed *BoneFinder* v.1.2.0, a software implementation provided by Lindner *et al.* [17], to segment the femoral scans using the CLM approach. All parameters were set as explained in [17].

2) Template Generation: General Procrustes analysis is adopted to find the reference template $\boldsymbol{T}$ [19]. First, all scans are aligned to a common position, scale, and orientation. Next, the reference template is updated as the average of the aligned shapes. The algorithm iterates between these two steps until convergence as detailed below.

The Procrustes analysis converges to a unique solution except for a scaling, rotation, or translation factor. To cancel out the arbitrary scaling of the template, the converged template was normalised with the scale $k = \left[\frac{1}{N}\left(k_1^* + \cdots + k_N^*\right)\right]^{-1}$ where k_n^* is the final scale factor after convergence for each individual shape. To cancel out the arbitrary rotation of the template, the template was rotated such that the bottom cross-section at the femoral shaft is parallel to the horizontal axis. To cancel out the arbitrary translation, the centre of gravity, i.e. the average of all landmark points on the template, is shifted to the origin at the $[0, 0]^T$ coordinate.

3) Pairwise Registration: To eliminate morphological variation between scans, each individual scan is warped to the template domain using thin-plate spline (TPS) registration [20]. In this technique, a geometrical transformation is found such that the landmark points in the source image are exactly mapped to their corresponding landmark points in the reference template. Given the transformation, the whole image space is warped to the template domain and the intensity values are linearly interpolated.

C. Comparative Calibration

Systematic measurement differences occur between densitometers from different manufacturers [5]–[7]. Discussing the biological or technical reasons for this discrepancy is not the purpose of this study, but to provide a universal standardisation of BMD values. The first attempt at cross-calibration between DXA scanners, sponsored by the International DXA Standardisation Committee (IDSC), showed that measurements across different machines are highly correlated [5]. In this study 100 healthy women were measured on three different scanners, i.e. Norland XR26 Mark I1; Lunar DPXL; and Hologic QDR 2000, demonstrating a linear relationship between each pair of scanners. To avoid designating one machine as the gold standard, they proposed an ad-hoc method to measure *true* or *standard BMD* (sBMD) [5]. This ad-hoc method has a few problems, as detailed in [6]. Later, Lu *et al.* [7] proposed a fully statistical methodology to solve this problem known as *comparative calibration*.

Assume C systems are each used to measure the same characteristics on a common set of N subjects. Each system may not be consistent in the repeated measurements of the same patient resulting in a *within-patient* sampling variation. However, this variation is assumed to be consistent for different patients. Ignoring this sampling fluctuation, the mean of repeated measurements is deemed to be the *true* underlying value that is not directly observable. Furthermore, we assume that a linear relationship exists between each pair of systems given the true underlying measurements. Then, *comparative calibration* refers to the problem of simultaneous estimation of the pairwise relationships between these systems [7], [21].

Let the latent random variable X represent the underlying true value and the random variable Y^c represents the observed value measured on the machine c. Barnett [21] proposed a linear model for comparative calibration between the systems:

$$Y^{(c)} = a_c X + b_c + E^{(c)}, \qquad \text{for } c = 1, \cdots, C. \tag{2}$$

$E^{(c)} \sim \mathcal{N}(0, \sigma_c^2)$ represents the measurement noise for each system and $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ represents the distribution of the population. Given the observed measurements $\boldsymbol{y}_n = \left[y_n^1, \cdots, y_n^C\right]^T$ for the subject n, the objective is to estimate the model parameters $\{a_c, b_c, \sigma_c\}_{c=1}^C$. This model is overparametrised and to resolve this identifiability problem, it is common to take one system, e.g. c^*, as the reference. For this system, then, it is assumed that $a_{c^*} = 1$ and $b_{c^*} = 0$ [21]. Alternatively, Lu *et al.* [7] added two extra linear equations:

$$\frac{1}{C}\sum_c b_c = B_0 \quad \text{and} \quad \frac{1}{C}\sum_c a_c = A_0, \tag{3}$$

where B_0 and A_0 are two constants defined based on either hypothetical assumptions or phantom measurements. Barnett [21] presented the solution for $C = 3$ using second order moment estimates. Lu *et al.* [7] presented an expectation maximisation (EM) approach to estimate model parameters for $C > 3$. For $C = 2$, the problem is known as *Deming Regression* [22].

These techniques require a complete set of measurements $\boldsymbol{y}_n$ for each subject and may not be deployed when only one single measurement is available for each subject due to insufficient statistics [7]. Requiring multiple measurements of each subject on all machines is a seriously limiting factor in large multicentre studies [9]. Here we propose a novel technique called *quantile matching regression* to tackle this problem.

Quantile Matching Regression: The new technique is developed based on two assumptions: First, a unique distribution of the latent variable X exists independent of the measurement systems. Second, the SNR is sufficiently large such that

$$Q_{Y^{(c)}}(u) \approx a_c Q_X(u) + b_c, \tag{4}$$

where $Q_X(u)$ and $Q_{Y^{(c)}}(u)$ denote the quantile functions. For a random variable X, the quantile function $u \rightarrow Q_X(u)$ is defined as $Q_X(u) := \inf\{x : u \leq \mathcal{P}(X \leq x)\}$. Therefore, if the noise power is zero, then the approximation would be replaced with equality in Eq. 4. With this assumption, estimation of the model parameters $\Theta = \{a_c, b_c\}$ can be decoupled from the estimation of noise variances, i.e. $\{\sigma_c^2\}$. This technique cannot estimate the noise variances because of insufficient statistics due to missing multiple measurements. However, this technique can provide reliable estimations for the slope a_c and intercepts b_c as detailed below.

The parameters Θ are estimated by minimising

$$\mathcal{J} = \frac{1}{2}\sum_{c=1}^{C}\int_0^1 (Q_{Y^{(c)}}(u) - a_c Q_X(u) - b_c)^2 du,$$
$$\text{subject to } \sum_c b_c = 0 \quad \text{and} \quad \frac{1}{C}\sum_c a_c = 1. \tag{5}$$

To set the constants in Eq. 3, we assume that the true value X equals the average of the expected observations given the latent variable X, i.e. $X = \frac{1}{C}\sum_c E(Y^{(c)}|X)$. This results in $B_0 = 0$ and $A_0 = 1$.

Optimisation: To convert the constrained optimisation problem into an unconstrained one, we can simply express the

Algorithm 1 Quantile Matching Regression

Input: $\mathcal{Y} = \{y_n^{(c_n)}\}_{n=1}^{N}$
Parameters: ϵ ▷ Convergence tolerance
Output: $\Theta = \{a_c, b_c\}_{c=1}^{C}$
procedure QUANTILE-MATCHING($\mathcal{Y}$)
  for $c = 1 : C$ **do**
    $Q_{Y^{(c)}}(u) \leftarrow$ Estimate quantile values for $\mathcal{Y}^c = \{y_n^{(c_n)} : c_n = c\}$
  $i \leftarrow 1$ ▷ Number of iterations
  $a_c \leftarrow 1$ and $b_c \leftarrow 0$ for $c = 1, \cdots, C$.
  while $\sum_{c=1}^{C}(a_c^{(i)} - a_c^{(i-1)})^2 + (b_c^{(i)} - b_c^{(i-1)})^2 \leq \epsilon^2$ **do**
    $i \leftarrow i + 1$
    for $n = 1 : N$ **do**
      $x_n \leftarrow E(X|y_n^{(c_n)}; \Theta_i)$ (Eq. 7)
    $Q_X(u) \leftarrow$ Estimate quantile values for $\mathcal{X} = \{x_1, \cdots, x_N\}$ ▷ (step 1)
    $a_c^{(i)}, b_c^{(i)} \leftarrow \underset{a_c, b_c}{\operatorname{argmin}} \frac{1}{2}\sum_c \int_0^1 (Q_{Y^{(c)}}(u) - a_c Q_X(u) - b_c)^2 du$ ▷ (step 2)

Fig. 4. Quantile matching regression technique for comparative calibration between C systems (Eq. 5). For each system, the quantile curve $Q_{Y(c)}$ is estimated using samples collected on the same system. Next, calibration parameters a_c and b_c are iteratively estimated by alternating between two steps: step one, estimate the standard quantile curve Q_X using samples collected on all systems, and step two, minimise the sum of difference between Q_X and each system-specific quantile curve $Q_{Y(c)}$.

parameters a_C and b_C based on the other parameters:

$$a_C = C - \sum_{c \neq C} a_c \quad \text{and} \quad b_C = -\sum_{c \neq C} b_c \tag{6}$$

To estimate the parameters, an alternating minimisation technique is adopted (Algorithm 1): Given the model parameters, the latent variable x_n for each of N subjects can be estimated as (step 1),

$$x_n = E(X|y_n^{(c_n)}; a_{c_n}, b_{c_n}) \approx \frac{1}{a_{c_n}}(y_n^{(c_n)} - b_{c_n}), \tag{7}$$

where c_n is the corresponding system for subject n. To update the model parameters, the gradients $\frac{\partial \mathcal{J}}{\partial a_c}$ and $\frac{\partial \mathcal{J}}{\partial b_c}$ are set to zero.

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial a_c} &= (a_c + \sum_{c' \neq C} a_{c'} - C)\int_0^1 Q_X(u)^2\, du \\ &\quad + (b_c + \sum_{c' \neq C} b_{c'})\int_0^1 Q_X(u) du \\ &\quad + \int_0^1 Q_X(u)(Q_{Y^{(C)}}(u) - Q_{Y^{(c)}}(u)) du = 0, \end{aligned} \tag{8}$$

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial b_c} &= (a_c + \sum_{c' \neq C} a_{c'} - C)\int_0^1 Q_X(u) du + (b_c + \sum_{c' \neq C} b_{c'}) \\ &\quad + \int_0^1 (Q_{Y^{(C)}}(u) - Q_{Y^{(c)}}(u)) du = 0. \end{aligned} \tag{9}$$

Computing $Q_X(u)$ using the estimated latent variables, Eq. 8 and Eq. 9 are linear with respect to the model parameters; we have $2(C-1)$ linear equations with $2(C-1)$ parameters for which a closed-form solution exists (step 2). The algorithm iterates between these two steps until the root mean square of the difference between estimated parameters at two consecutive iterations is less than a user-defined tolerance ϵ.

D. Quantile Regression

Assume the real-valued random variable Y with *cumulative distribution function* (CDF) $F_Y(y) = P(Y \leq y)$ represents a response variable of interest, e.g. BMD values at a single pixel coordinate, and the multivariate random variable $\boldsymbol{X} = [X_1, \cdots, X_p]^T$ represents an explanatory covariate vector, e.g. age, BMI, etc. Then, the conditional quantile function $(u, \boldsymbol{x}) \longmapsto Q_{Y|X=\boldsymbol{x}}(u, \boldsymbol{x})$ is defined as

$$Q_{Y|X}(u, \boldsymbol{x}) := \inf\left\{y : u \leq F_{Y|X=\boldsymbol{x}}(y)\right\}, \tag{10}$$

where $0 < u < 1$. The main objective is to estimate $Q_{Y|X}(u, \boldsymbol{x})$ from N observed scattered points $(y_n, \boldsymbol{x}_n)$.

In [13], we formulated the problem for a scalar covariate, i.e. age. Here, we extend the methodology to include other covariates of interest, e.g. BMI, using the vector generalised additive models (VGAMs) [23]. VGAMs model the conditional probability distribution $\mathcal{P}(y|\boldsymbol{x}) = h(y, \eta_1, \cdots, \eta_M)$, where $h(.)$ is a known function and $\boldsymbol{\eta} = [\eta_1, \cdots, \eta_M]^T$ are *linear predictors*. The m^{th} predictor η_m is estimated as the sum of smooth functions of the individual covariates x_p,

$$\eta_m = \eta_m(\boldsymbol{x}) = f_{(m)0} + \sum_{p=1}^{P} f_{(m)p}(x_p). \tag{11}$$

To parametrise the model, we use the LMS technique [24] and thereby $M = 3$, $\eta_1 = \lambda$, $\eta_2 = \mu$, and $\eta_3 = \sigma$. The LMS method assumes a Box-Cox transformation (Eq. 12) with appropriate parameters exists such that the positive random variable Y can be mapped to a standard normal distribution $Z \sim \mathcal{N}(0, 1)$.

$$Z = \begin{cases} \dfrac{(\frac{Y}{\mu(\boldsymbol{x})})^{\lambda(\boldsymbol{x})} - 1}{\sigma(\boldsymbol{x})\lambda(\boldsymbol{x})}, & \lambda(\boldsymbol{x}) \neq 0; \\ \dfrac{1}{\sigma(\boldsymbol{x})} \ln(\dfrac{Y}{\mu(\boldsymbol{x})}), & \lambda(\boldsymbol{x}) = 0. \end{cases} \tag{12}$$

Under the LMS normality assumption, the log-likelihood of the parameters is

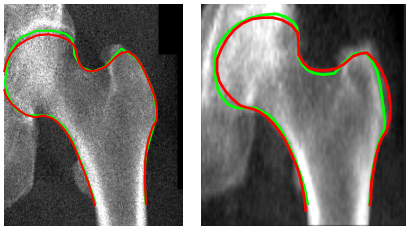
$$\ell = \sum_{n=1}^{N}\left[\lambda(\boldsymbol{x}_n)\ln\frac{y_n}{\mu(\boldsymbol{x}_n)} - \ln\sigma(\boldsymbol{x}_n) - \frac{1}{2}z_n^2\right]. \tag{13}$$

To estimate smooth functions $f_{(m)p}(x)$, we deployed the R-package VGAM to maximise the penalised likelihood below [23].

$$\mathcal{J} = \ell - \frac{1}{2}\sum_{p=1}^{P}\sum_{m=1}^{M}\alpha_{(m)p}\int f''_{(m)p}(x_p)^2 dx_p. \tag{14}$$

Estimating smooth functions $\lambda(\boldsymbol{x})$, $\mu(\boldsymbol{x})$, and $\sigma(\boldsymbol{x})$, quantiles can be simply computed;

$$Q_{Y|X}(u, \boldsymbol{x}) = \begin{cases} \mu(\boldsymbol{x})\left[1 + \lambda(\boldsymbol{x})\sigma(\boldsymbol{x})Q_Z(u)\right]^{1/\lambda(\boldsymbol{x})}, & \lambda(\boldsymbol{x}) \neq 0; \\ \mu(\boldsymbol{x})\exp\left(\sigma(\boldsymbol{x})Q_Z(u)\right), & \lambda(\boldsymbol{x}) = 0. \end{cases} \tag{15}$$



(a) DSC = 0.9801 (b) DSC = 0.9620

Fig. 5. The best and the worst femoral segmentation in the test dataset ($n = 32$). The green and the red contours show the ground truth and the automatic segmentation, respectively.

In our experiments, we have modelled the two parameters λ and σ as intercepts. To control the smoothness on the parameter μ, the equivalent degree of freedom (edf) was set to 3. The optimisation procedure is numerically complex and the algorithm failed to converge for a fraction of pixels ($\approx 1.3\%$). For these pixels, the outliers are first removed and then the algorithm was run on the cleaned data.

To assess the precision of the estimated quantile curves, a bootstrapping procedure was deployed; subjects were randomly sampled with replacement and the quantiles were re-estimated. This procedure was repeated $1,000$ times, collecting a distribution of possible quantile values. From these observations, the confidence intervals at 5% significance level were estimated [25].

III. Results and Experiments

A. Data

To generate the spatio-temporal bone ageing atlas over the adulthood (20-97 years), we integrated data from three Western European population studies: The UK Biobank [26] ($n = 6,918$, age $= 45 - 80$ years, White women), the Osteoporosis and Ultrasound study (OPUS) [27] ($n = 1,402$; age $= 20 - 39$ and $55 - 79$ years; White women), and the MRC-Hip study [28] ($n = 5,018$; age $= 75 - 97$ years; White women). Scans were collected using either an iDXA Lunar GE (the UK Biobank study) or a Hologic QDR4500 Acclaim densitometer (the OPUS and MRC-Hip study cohorts).

B. Segmentation Accuracy

To evaluate the segmentation accuracy, a subset of scans ($n = 32$) randomly selected from the database were manually annotated. The segmentation accuracy was evaluated using the Dice similarity coefficient (DSC). DSC is defined as the twice the areal size of the overlap between two binary masks divided by the sum of the areal size of each mask. The mean and the standard deviation for DSC over the 32 selected scans were 0.9698 and 0.0048, respectively. Fig. 5 shows the worst and the best segmentation results based on the DSC metric. Since the cut-off point at the femoral shaft is arbitrary, the shorter distal cut-off point between the manual and the automated masks is used to cancel out the variation in the shaft before computing the DSC metric.

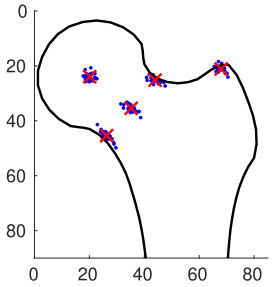


Fig. 6. Point localisation error. Five landmark points (blue dots) were selected manually at anatomically correspondent locations and then mapped to the reference domain using the estimated TPS transformations for each image ($n = 32$). The average error was 1.57 mm. The space is shown in millimetre.

C. Point Localisation Accuracy

The same dataset ($n = 32$) used for the evaluation of segmentation accuracy was deployed here. To evaluate the point localisation accuracy, five landmark points were selected manually at key prominent geometrical locations: centre of the femoral head; the centre, superior, and inferior positions at the femoral neck; and, finally, the apex at the greater trochanter. The landmarks were then transferred to the template using the same TPS warping transformation computed per each image (Fig. 6). The overall error was 1.57 mm [29].

D. Precision Analysis

Precision or reproducibility of a quantitative measurement technique describes its ability to produce consistent results when measuring the same quantity repeatedly. In other words, precision is a description of random errors in the system. Three sources of error exist [30]: the machine (e.g., the scanner noise), the operator (e.g., patient positioning), and the software (e.g., femur segmentation and deformable image alignment).

To assess the overall precision of the RFA technique, 25 Caucasian women (mean age $= 70.1 \pm 6.2$ years) were scanned on the same day twice with repositioning between the scans. This data had been collected as part of the OPUS study in Sheffield. In conventional DXA analysis, precision is reported as the coefficient of variation (CV), i.e. the root mean square standard deviation divided by the mean of paired measurements, for the selected ROIs [31].

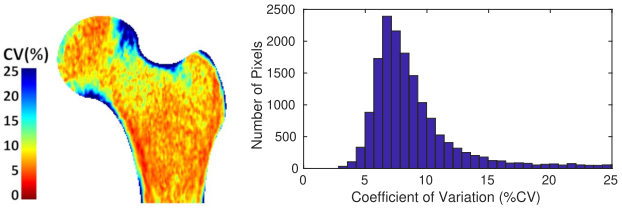
$$CV = 100\% \times \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{(y_n - y'_n)^2}{2}}}{\frac{1}{N}\sum_{n=1}^{N}\frac{(y_n + y'_n)}{2}} \tag{16}$$

Here, $N = 25$ is the number of paired measurements; y and y' are the measured BMD values at the two independent positions.

Table I reports the precision of conventional region-based DXA analysis at four common ROIs. To use RFA to reproduce conventional region-based analysis, pixel BMD values of the warped scans were aggregated at each ROI in the template domain. RFA resulted in similar precision scores to those reported in the literature at these ROIs (Table I). However, as anticipated, a finer pixel-level analysis using the RFA technique results in poorer precision. Fig. 7(a) shows the distribution of pixel-level CV values at the proximal femur.

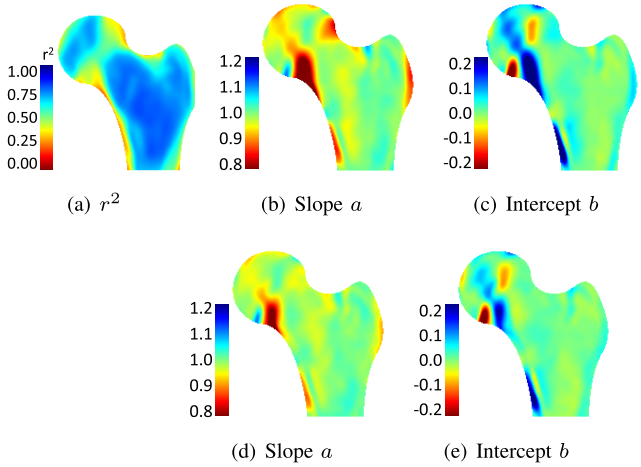
TABLE I
COEFFICIENT OF VARIATION (%) AT FOUR COMMON ROIS

method	scanner	subjects No. × scans No.	CV% total hip	neck	trochanter	intertrochanter
RFA	Hologic QDR 4500A	25 × 2	1.05	1.73	1.87	1.14
[32]	Hologic QDR 2000	71 × 2	1.2	1.7	1.4	1.7
[33]	Hologic QDR 4500A	27 × 2	1.69	1.11	1.27	-
[34]	Lunar Prodigy	6 × 6	0.65	1.66	1.16	-
[31]	Hologic QDR 2000	95 × 2	1.59	2.25	-	-



(a) Heat-map (b) Histogram

Fig. 7. DXA RFA precision analysis. (a) The pixel level CV (%) is visualised using a heat-map. Precision is worse around the bone contour. This may be due to the inaccuracy in placing controlling landmark points at the bone surface. (b) The distribution of pixel-level CV values in the femur. The median is 7.96% and the interquartile range is 6.69%−10.05%.



(a) r^2 (b) Slope a (c) Intercept b

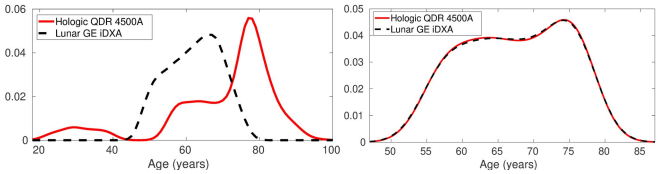
(d) Slope a (e) Intercept b

Fig. 8. Bilateral hip calibration. (a) The left and the right hips are highly correlated inside the femur, but the correlation is worse at the boundary. Estimated cross-calibration parameters between the left and the right hips, i.e. [left] = a [right] + b, are shown for the Deming regression technique in panels (b) and (c); and for the quantile matching regression technique in panels (d) and (e).

Precision was worse around the bone contours. This may be explained due to the inaccuracy in placing controlling landmark points around the bone. Fig. 7(b) shows the histogram of pixel-level CV values where the median is 7.96% and the interquartile range is 6.69% − 10.05%. The worse precision in comparison to conventional region-based analysis is a compromise that is offset against the substantially finer spatial analysis that is necessary for characterising spatially complex bone remodelling events.

E. Comparative Calibration

1) Validation: Since no paired measurements acquired on both the Hologic and the Lunar systems were available for analysis, it was not possible to test the viability of the proposed quantile matching regression technique directly for DXA cross-calibration. Alternatively, we tested the technique in the setting of bilateral hip calibration using $n = 6,916$ DXA measurements from both left and right hips as part of the UK Biobank study. Given the good correlation between BMD of the left and the right hips (8(a)), one can postulate that bilateral BMD values are noisy measurements of the same underlying hidden variable except for a linear transformation (Eq. 2). Since we had access to paired measurements for each subject, Deming regression would give the ground truth for the calibrations parameters (Figs. 8(b) and 8(c)). Ignoring the fact bilateral hip scans came from the same subject, quantile matching regression was deployed to approximate the calibration parameters (Figs. 8(d) and 8(e)). The estimated parameters are similar to those computed using Deming regression. Over the region with a high correlation between the left and the



(a) (b)

Fig. 9. Sample selection for cross-calibration between Hologic QDR4500A and iDXA Lunar GE scanners. (a) The probability distribution of the age for white British women recruited in this study before sample selection. (b) The probability distribution of the age for $n = 406$ subjectes matched for age and BMI to be used in quantile matching regression. No significant difference was observed between the age distributions following sample section using a two-sample Kolmogorov-Smirnov test (p-value = 0.9).

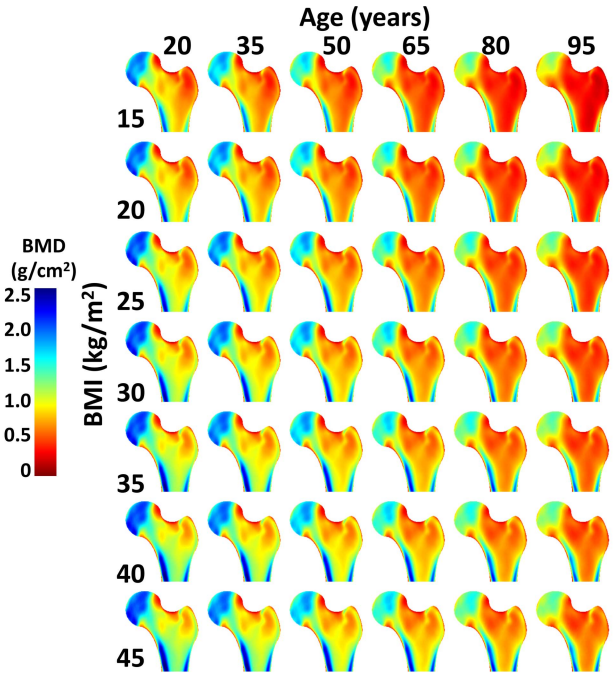
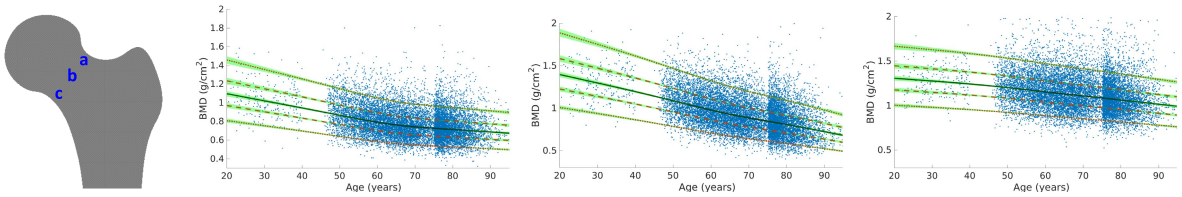
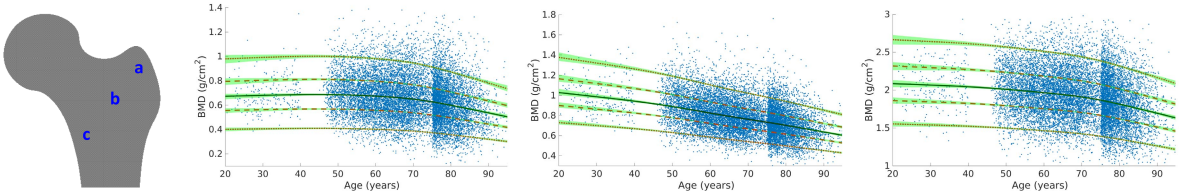


Fig. 10. Pixel-level median BMD values are visualised using heat-maps as a function of age for 20, 35, 50, 65, 80, and 95 years and BMI values of 15, 20, 25, 30, 35, 40, and 45 kg/m^2. The atlas is shown for the Hologic system at the left hip.



(a) Superior Femoral Neck (b) Mid Femoral Neck (c) Inferior Femoral Neck

Fig. 11. Three examples of fitted quantile curves at three different pixel locations at the femoral neck region. The solid, dashed, and dotted lines show the median, 50% and 90% quantile ranges, respectively. The green shadow shows the 95% confidence interval. The curves are shown for the Lunar system at the left hip at median BMI = 25.4 kg/m^2.



(a) Superior Trochanteric Region (b) Mid Trochanteric Region (c) Inferior Trochanteric Region

Fig. 12. Three examples of fitted quantile curves at three different pixel locations at the intertrochanteric region. The solid, dashed, and dotted lines show the median, 50% and 90% quantile ranges, respectively. The green shadow shows the 95% confidence interval. The curves are shown for the Lunar system at the left hip at median BMI = 25.4 kg/m^2.

right hips ($r^2 \geq 0.5$), the RMS error was 0.013 for the slope a and 0.017 for the intercept b, respectively

2) DXA Cross-Calibration: In this study, scans were collected either on a Hologic QDR 4500A or an iDXA Lunar GE scanner. For each scanner, $n = 406$ white British women matched for age and BMI with an scan on the left side were selected. Fig. 9 shows the age distribution of subjects before and after this sample selection step. No significant difference in age or BMI distribution was observed between the two groups using a two-sample Kolmogorov-Smirnov test (p-value = 0.9). Note that although the calibration factors depend only on the technical properties specific to each imaging system, our proposed quantile matching regression still requires age- and BMI-normalisation to ensure any variation in BMD distributions is due only to the difference in imaging system technologies rather than patient characteristics. The average and standard deviation of the estimated parameters over all pixels within the femur were 1.019 (SD, 0.140) for the slope a and 0.170 (SD, 0.130) for the intercept b.

We conducted a further validation experiment to confirm the independence of estimated calibration parameters from cohort age. Out of 406 subject pairs matched for age and BMI, subjects from two age segments 60-65 years and 70-75 years ($n = 148$) were left out for testing and the remaining scans (n=258) were deployed for estimation of calibration parameters. Next, we tested the validity of the fitted parameters on the test data using a two-sample Kolmogorov-Smirnov test with FDR correction. No significant differences were observed between the testing versus the training datasets, confirming that the calibration parameters are extendable to age segments 60-65 and 70-75 years.

F. The Spatio-Temporal Atlas

Fig. 10 shows the constructed atlas; visualising median BMD values at different values of age and BMI using heat-maps. Low BMI was associated with an overall decrease in bone mass, where as high BMI resulted in increased bone mass especially at the diaphysis and Ward's triangle regions. An overall decline in BMD with increasing age was observed throughout the proximal femur. However, the observed bone loss patterns were site-specific and spatially-complex. Cortical thinning was observed consistently with ageing around the femoral shaft from the 6^{th} decade onwards. Widespread bone loss was also observed in the trochanteric area.

Quantile regression curves demonstrated different rates of bone loss at different anatomic locations within the proximal femur (Figs. 11 and 12). For example, the decrease in BMD at the superior femoral neck cortex was bimodal; the bone loss slowed down from the 70s onwards (Fig. 11(a)). BMD at the mid-femoral neck showed a steady decrease throughout the whole age range (Fig. 11(b)), whilst bone mass was preserved the most in the inferior femoral neck cortex (Fig. 11(c)). Fig. 12 shows quantile regression curves at the intertrochanteric region. Bone mass at the superior trochanteric region was preserved until just before 70 years, and was followed by a decline with a similar slope to the other trochanteric regions (Fig. 12(a)). Bone loss was observed at a consistent rate at the mid trochanteric region throughout the whole age range (Fig. 12(b)). BMD in the inferior cortex close to the lesser trochanter was maintained until the 60^{th} year, following which point BMD showed a steady decline (Fig. 12(c)).

The inflection point observed at age 75 in Fig. 12 is indeed due to ageing rather than the integration of the MRC-Hip dataset (age range: 75-97 years). Repeating the same analysis using only the UK Biobank dataset (age range: 45-80 years) demonstrated similar ageing trends (data not shown). Here, the results for the integration of all datasets together is presented.

G. Atlas Validation Using Longitudinal Data

The bone ageing atlas was developed based on cross-sectional data. We acknowledge that this atlas does not

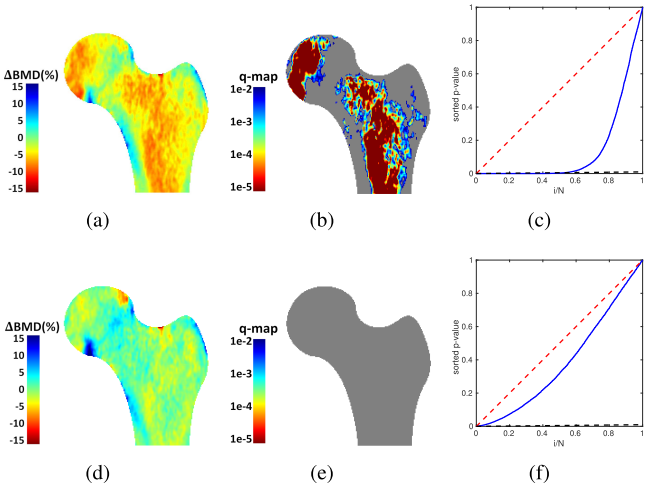


Fig. 13. Atlas validation using paired longitudinal data collected at baseline and 6 years later ($n = 120$). (a) Normalised BMD change between baseline and follow-up measurements at six years. (b) FDR q-map corresponding to panel a indicating regions with a significant BMD change at 6 years. (c) The PP-plot corresponding to panel b. If the null hypothesis of no significant BMD change is true, then the blue line follows the identity (dashed red line). (d) Normalised BMD difference between baseline maps projected at 6 years and the actual follow-up measurements at 6 years. (e) FDR q-map corresponding to panel d. (f) The PP-plot corresponding to panel e. These data show that the BMD change projected by the atlas and the directly observed BMD change are quantitatively and statistically similar, confirming the viability of the developed atlas to predict temporal BMD change.

necessarily provide an ideal prediction of individual ageing, for which, longitudinal data with the same subjects repeatedly scanned along several years is required. However, the utility of the developed atlas to predict longitudinal changes is tested here using a subset of scans from the OPUS dataset ($n = 120$; age range=55-60 years) with follow-up measurements at 6 years (mean time lapse, 70.9 months; standard deviation, 1.2 months). The hypothesis tested here is that no significant BMD change should be observed between the expected BMD values at 6 years based on the projected BMD atlas and the actual measurements at 6 years. For this analysis, a paired t-test preceded by false discovery rate (FDR) analysis [11] was used once between the baseline and the actual follow-up measurements, and another time between the projected and the actual follow-up BMD values.

To project BMD values six years into the future, firstly, the quantile value for the given pixel BMD at the baseline age is read from the atlas. Next, the corresponding BMD value at the follow-up age is read from the same quantile trajectory. Significant bone loss was observed in the trochanteric region and the medial femoral shaft; however, the projected BMD values using the constructed atlas fits the actual measurements where no significant BMD change was observed between the projected and the actual BMD values (Fig. 13).

IV. Conclusion

This work presents the development of a reference spatio-temporal atlas of ageing bone in the proximal femur using cross-sectional data from a large cohort of Western European Caucasian women (n=13,338). Here, we presented three key contributions: first, the proposed DXA RFA framework allowed high-resolution pixel-level BMD analysis. The increased spatial detail made it possible to observe spatially-complex bone ageing patterns for which conventional region-based bone densitometry routine is insensitive. Second, the proposed calibration technique allowed the integration of data from different DXA manufacturers. The new method does not require multiple scans from the same subject and so is applicable to large multi-centre studies. Third, a fully automatic bone ageing analysis pipeline was proposed that would streamline the atlas generation process. This automation would facilitate population-specific atlas generation from other ethnic libraries.

Each module in the pipeline was evaluated separately. The average segmentation accuracy expressed as the Dice index was 0.97. The average point localisation error was 1.57 mm equivalent to 3.15 pixels. The RFA precision expressed as median pixel-level coefficient of variation was 7.96%. The RMS error for quantile matching regression tested in the setting of the bilateral hip calibration was 0.013 and 0.017 for the slope and the intercept parameters, respectively. The precision of the LMS quantile regression for modelling the temporal BMD evolution was tested using a bootstrapping procedure. The overall uncertainty was sufficiently small so the ageing effect was observable (Figs. 11 and 12). We demonstrated the utility of the proposed bone ageing analysis pipeline using three large-scale datasets with $n > 13,000$ scans collected on two different manufacturer's densitometers. However, the proposed pipeline would facilitate population-specific atlas generation from other ethnic libraries, gender, and anatomic sites. This, in turn, would allow the analysis of variations in ageing patterns across different populations.

This technique also had limitations. The areal BMD measured by DXA does not represent the true volumetric BMD, and so the constructed atlas is a 2D projection of the actual 3D patterns. A 2D/3D approach could address this issue [35], [36]. These techniques are often based a 3D statistical shape/appearance model learned from a small subset of QCT images, for example, $n = 57$ (all highly osteoporotic women) [35]. Hence, the learned atlas cannot account for the full population variation (cf. $n = 13,338$ in this study). If a large QCT dataset was available, the ageing atlas could have been directly developed from them where the principle applied here can be readily transferred to 3D imaging.

This technique shows promise in characterising spatially-complex BMD changes with ageing. These patterns were visualised using heat-maps. Furthermore, quantile curves plotted at different pixel coordinates showed consistently different rates of bone loss at different regions of the femoral neck. Our future work aims at improving fracture risk assessment using the developed atlas to determine whether this increased resolution enhances the fracture predictive ability of DXA.

Acknowledgment

This work was undertaken using the UK Biobank Resource (AA#17881), and data from the MRC-Hip study (MRC/G9518113) and OPUS study (sponsored by Eli Lilly, Sanofi-Aventis, Procter & Gamble Pharmaceuticals,

Hoffman-La Roche, Pfizer, and Novartis). The authors thank the OPUS Steering Committee for permission to use its DXA scans.

REFERENCES

[1] S. Khosla and B. Riggs, "Pathophysiology of age-related bone loss and osteoporosis," *Endocrinol. Metabolism Clinics*, vol. 34, no. 4, pp. 1015–1030, 2005.

[2] O. Demontiero, C. Vidal, and G. Duque, "Aging and bone loss: New insights for the clinician," *Therapeutic Adv. Musculoskeletal Disease*, vol. 4, no. 2, pp. 61–76, 2012.

[3] W. Huizinga *et al.*, "A spatio-temporal reference model of the aging brain," *NeuroImage*, vol. 169, pp. 11–22, Apr. 2017.

[4] J. E. Adams, "Advances in bone imaging for osteoporosis," *Nature Rev. Endocrinol.*, vol. 9, no. 1, pp. 28–42, 2013.

[5] H. Genant *et al.*, "Universal standardization for dual X-ray absorptiometry: Patient and phantom cross-calibration results," *J. Bone Mineral Res.*, vol. 9, no. 10, pp. 1503–1514, 1994.

[6] S. Hui *et al.*, "Universal standardization of bone density measurements: A method with optimal properties for calibration among several instruments," *J. Bone Mineral Res.*, vol. 12, no. 9, pp. 1463–1470, 1997.

[7] Y. Lu, K. Ye, A. K. Mathur, S. Hui, T. Fuerst, and H. K. Genant, "Comparative calibration without a gold standard," *Statist. Med.*, vol. 16, no. 16, pp. 1889–1905, 1997.

[8] J. Pearson *et al.*, "European semi-anthropomorphic spine phantom for the calibration of bone densitometers: Assessment of precision, stability and accuracy the European quantitation of osteoporosis study group," *Osteoporosis Int.*, vol. 5, no. 3, pp. 174–184, 1995.

[9] D. Reid *et al.*, "Cross-calibration of dual-energy X-ray densitometers for a large, multi-center genetic study of osteoporosis," *Osteoporosis Int.*, vol. 17, no. 1, pp. 125–132, 2006.

[10] R. Morris, L. Yang, M. A. Martín-Fernández, J. M. Pozo, A. F. Frangi, and J. M. Wilkinson, "High-spatial-resolution bone densitometry with dual-energy X-ray absorptiometric region-free analysis," *Radiology*, vol. 274, no. 2, pp. 532–539, 2014.

[11] M. Farzi *et al.*, "Quantitating the effect of prosthesis design on femoral remodeling using high-resolution region-free densitometric analysis (DXA-RFA)," *J. Orthopaedic Res.*, vol. 35, no. 10, pp. 2203–2210, 2017.

[12] A. M. Parker, L. Yang, M. Farzi, J. M. Pozo, A. F. Frangi, and J. M. Wilkinson, "Quantifying pelvic periprosthetic bone remodeling using dual-energy X-ray absorptiometry region-free analysis," *J. Clin. Densitometry*, vol. 20, no. 4, pp. 480–485, 2017.

[13] M. Farzi, J. M. Pozo, E. McCloskey, R. Eastell, J. M. Wilkinson, and A. F. Frangi, "Spatio-temporal atlas of bone mineral density ageing," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 720–728.

[14] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models," *Med. Image Anal.*, vol. 6, no. 1, pp. 47–62, 2002.

[15] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognit.*, vol. 41, no. 10, pp. 3054–3067, 2008.

[16] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for medical image analysis and computer vision," *Proc. SPIE*, vol. 4322, pp. 236–248, Jul. 2001.

[17] C. Lindner, S. Thiagarajah, J. M. Wilkinson, T. A. Consortium, G. A. Wallis, and T. F. Cootes, "Fully automatic segmentation of the proximal femur using random forest regression voting," *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1462–1472, Aug. 2013.

[18] J. Gall and V. Lempitsky, "Class-specific Hough forests for object detection," in *Decision Forests for Computer Vision and Medical Image Analysis*. London, U.K.: Springer, 2013, pp. 143–157.

[19] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Roy. Statist. Soc.*, vol. 53, no. 2, pp. 285–339, 1991.

[20] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.

[21] V. D. Barnett, "Simultaneous pairwise linear structural relationships," *Biometrics*, vol. 25, no. 1, pp. 129–142, 1969.

[22] W. E. Deming, *Statistical Adjustment of Data*. Hoboken, NJ, USA: Wiley, 1943.

[23] T. W. Yee, "Quantile regression via vector generalized additive models," *Statist. Med.*, vol. 23, no. 14, pp. 2295–2315, 2004.

[24] T. J. Cole and P. J. Green, "Smoothing reference centile curves: The LMS method and penalized likelihood," *Statist. Med.*, vol. 11, no. 10, pp. 1305–1319, 1992.

[25] J. Carpenter and J. Bithell, "Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians," *Statist. Med.*, vol. 19, no. 9, pp. 1141–1164, 2000.

[26] C. Sudlow *et al.*, "UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS Med.*, vol. 12, no. 3, 2015, Art. no. e1001779.

[27] C. Glüer *et al.*, "Association of five quantitative ultrasound devices and bone densitometry with osteoporotic vertebral fractures in a population-based sample: The OPUS study," *J. Bone Mineral Res.*, vol. 19, no. 5, pp. 782–793, May 2004.

[28] E. McCloskey *et al.*, "Clodronate reduces the incidence of fractures in community-dwelling elderly women unselected for osteoporosis: Results of a double-blind, placebo-controlled randomized study," *J. Bone Mineral Res.*, vol. 22, no. 1, pp. 135–141, 2007.

[29] M. Schröder, H. Gottschling, N. Reimers, M. Hauschild, and R. Burgkart, "Automated morphometric analysis of the femur on large anatomical databases with highly accurate correspondence detection," *Open Med. J.*, vol. 1, no. 1, pp. 15–22, 2014.

[30] S. Baim, C. R. Wilson, E. M. Lewiecki, M. M. Luckey, R. W. Downs, Jr., and B. Lentle, "Precision assessment and radiation safety for dual-energy X-ray absorptiometry: Position paper of the international society for clinical densitometry," *J. Clin. Densitometry*, vol. 8, no. 4, pp. 371–378, 2005.

[31] M. Lodder *et al.*, "Reproducibility of bone mineral density measurement in daily practice," *Ann. Rheumatic Diseases*, vol. 63, no. 3, pp. 285–289, 2004.

[32] S. Henzell *et al.*, "Precision error of fan-beam dual X-ray absorptiometry scans at spine, hip, and forearm," *J. Clin. Densitometry*, vol. 3, no. 4, pp. 359–364, 2000.

[33] S. L. Morgan, W. Abercrombie, and J. Y. Lee, "Need for precision studies at individual institutions and assessment of size of regions of interest on serial DXA scans," *J. Clin. Densitometry*, vol. 6, no. 2, pp. 97–101, 2003.

[34] J. White, S. S. Harris, G. E. Dallal, and B. Dawson-Hughes, "Precision of single vs bilateral hip bone mineral density scans," *J. Clin. Densitometry*, vol. 6, no. 2, pp. 159–162, 2003.

[35] O. M. Ahmad, K. Ramamurthi, K. E. Wilson, K. Engelke, M. Bouxsein, and R. H. Taylor, "3D structural measurements of the proximal femur from 2D DXA images using a statistical atlas," *Proc. SPIE*, vol. 7260, Feb. 2009, Art. no. 726005.

[36] T. Whitmarsh, L. Humbert, M. De Craene, L. M. Del Río Barquero, and A. F. Frangi, "Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry," *IEEE Trans. Med. Imag.*, vol. 30, no. 12, pp. 2101–2114, Dec. 2011.